Reading Group Notes

Joshua Agterberg

June 16, 2020

1 Notation

Throughout, I will use mostly the same notation as in Cape et al. (2019b). The main difference is I will only examine the results as they pertain to rank r matrices. We write $A = P + E$, where $P = U\Lambda U^\top$ is rank r and $A = \hat{U}\hat{\Lambda}\hat{U}^\top + \hat{U}_\perp\hat{\Lambda}_\perp\hat{U}_\perp^\top$.

For a concrete example to keep in mind and to focus our study on random graphs, I will specify to the setting wherein

$$P = \rho Z B Z^\top = \rho \left(\begin{array}{cccc|cccc} a & a & \cdots & a & b & b & \cdots & b \\ a & a & \cdots & a & b & b & \cdots & b \\ \vdots & \vdots & \ddots & \vdots & \vdots & \vdots & \vdots & \vdots \\ a & a & \cdots & a & b & b & \cdots & b \\ \hline b & b & \cdots & b & a & a & \cdots & a \\ \vdots & \vdots & \ddots & \vdots & \vdots & \vdots & \vdots & \vdots \\ b & b & \cdots & b & a & a & \cdots & a \\ b & b & \cdots & b & a & a & \cdots & a \end{array}\right).$$

Here the matrix is rank 2 with eigenvector matrix U defined as

$$U = \begin{pmatrix} \frac{1}{\sqrt{n}} & \frac{1}{\sqrt{n}} \\ \vdots & \vdots \\ \frac{1}{\sqrt{n}} & \frac{1}{\sqrt{n}} \\ \frac{1}{\sqrt{n}} & \frac{-1}{\sqrt{n}} \\ \vdots & \vdots \\ \frac{1}{\sqrt{n}} & \frac{-1}{\sqrt{n}} \end{pmatrix}.$$

The nonzero eigenvalues of P are $\frac{n\rho(a+b)}{2}$ and $\frac{n\rho(a-b)}{2}$.

2 Notes on Cape et al. (2019b)

The high-level overview is that the paper presents a series of matrix perturbation results for controlling the entrywise error from $\hat{U}$ to U. The main theorems are particularly user-friendly in the low-rank setting (e.g. the one above) and can be adapted a bit in the more general setting. I think of this paper as the paper that really emphasized and developed nice techniques using the $2 \to \infty$ norm to get good entrywise guarantees. The results therein are not necessarily the tightest bounds, but they are some of the most general bounds one can get without additional assumptions. If you want to do or use any entrywise eigenvector analysis, there are a number of results, tools, and definitions in this paper that are useful.

2.1 Matrix Perturbation Foundations

If you do not know about matrix perturbation theorems, I think one of the best ways to learn is a combination of working on matrix perturbation results, studying the end of Cai and Zhang (2018), and examining the results in Cape et al. (2019b). Bhatia is the best textbook on this in my opinion. Stewart and Sun is okay, and if you don't like the somewhat analytical nature of Bhatia, it's probablly the best option, but their presentation is a bit inscrutable I think.

If you don't know already, it's good to know all the ways you can write the $\sin\Theta$ distance, which is said somewhere in the paper. However, I do know that Cai and Zhang (2018) has these results proven explicitly, though they are not that hard. In particular,any of the terms

$$\begin{aligned} U^\top \hat{U} \\ \sin\Theta \\ (I - UU^\top)\hat{U} \\ UU^\top - \hat{U}\hat{U}^\top, \end{aligned}$$

in whatever orthogonally invariant norm you like should always scream $\sin\Theta$.

I think it is very important to understand the role of the orthogonal matrix W_*. When U is as in the example, W_* takes the form of sign-flips, which is why the results of Abbe et al. (2017) are reported the way they are (up to sign).

2.2 Entrywise Perturbation

However, if you want to get up and running on understanding entrywise subspace perturbation, I think the following things from Cape et al. (2019b) are useful.

- Lemma 6.7, in particular that

  $$||U^\top \hat{U} - W_*|| \leq ||\sin\Theta||_2^2.$$

  This is super useful with a quick additional Davis-Kahan application.

- All of section 6.1 with the properties of the $2 \to \infty$ norm, particularly Proposition 6.5 and the fact that $||UV||_{2,\infty} \leq ||U||_{2,\infty}||V||_2$.

- The beginning of the proof of Theorem 3.1 which writes

  $$\hat{U} - UW_* = \hat{U} - UU^\top \hat{U} - U[W - U^\top \hat{U}].$$

  The first term is much much easier to bound in $2 \to \infty$ norm (more on that later), and the second term, combined with lemma 6.7, is very easy to bound.

- Theorem 4.2. It is a nice, straightforward, and general bound on the $2 \to \infty$ difference. Whenever I get some result that doesn't neatly fit into something already studied, I use this bound as a check.

- Something that is sort of noted throughout the paper but not explicitly stated as a theorem: the dominant term is $EU\Lambda^{-1}$. If you know the CLT for the eigenvectors, you realize that this term is the big boi, and if you know the CLT for $\hat{X}$ for RDPGs (what Carey calls Avanti's CLT but Avanti does not call Avanti's CLT), you know that this term is the big boi by decreasing the power on Λ by a half.

- I call Theorem 3.1 and its associated corollaries examples of a "grand decomposition." My current working theory is that the tightest possible bounds can be achieved using the most intricate and careful grand decompositions and respective analyses. Abbe et al. (2017) has a grand decomposition very closely related, and Lei (2019) does as well, though Abbe et al. (2017) has only a term $A\hat{U}\Lambda^{-1}$ and Lei (2019) has the term $E\hat{U}\Lambda^{-1}$ which is easier to control. As far as I know, the techniques used in those papers are the state-of-the-art (leave-one-out analysis, kato integrals), though the techniques used therein are fundamentally much more involved then those used in Cape et al. (2019b) or even Cape et al. (2019a). I believe that all results on entrywise subspace perturbation can essentially be broken down into a) choice of grand decomposition and b) techniques used. Apparently Lei (2019) has the tightest bounds around.

2.3 Important Theorems

As discussed already, Theorem 4.2 a nice result for easy comparison.

Theorem 1 (Theorem 4.2)**.** *Let $A = P + E$ with rank$(P) = r$ and $P = U\Lambda U^\top$ with leading eigenvalues $|\lambda_1|, ..., |\lambda_r| > 0$. If $|\lambda_r| \geq 4||E||_\infty$, then there exists an orthogonal matrix W_U such that*

$$\|\hat{U} - UW_U\|_{2,\infty} \leq 14 \left(\frac{\|E\|_\infty}{|\lambda_r|} \right) \|U\|_{2,\infty}.$$

I see this (and the grand decomposition) as the big theorems in the paper. For some situations, the grand decomposition and the theorem above suffice to get a result (e.g. Zhu et al. (2019) use the grand decomposition Theorem 3.1 Cape et al. (2019b) in its proofs), but in other situations, one must start with some grand decomposition and modify existing techniques to get possibly tighter bounds. I think Little et al. (2019) also use the analogue of Theorem 4.2 in Abbe et al. (2017) for their CMDS result, but (I believe) their result is not tight – but it suffices for their particular analysis.

Theorem 4.2 is quite nice, but I also like Theorem 4.7.

Theorem 2 (Theorem 4.7, adapted for these notes)**.** *Let A_1 and A_2 be two ν-correlated SBM graphs. Suppose both graphs have $n/2$ vertices within each block, and suppose B is in our example above. Set*

$$A := \begin{pmatrix} A_1 & \frac{A_1+A_2}{2} \\ \frac{A_1+A_2}{2} & A_2 \end{pmatrix}.$$

Then if $n\rho_n \gg \log^4(n)$, there exists an orthogonal matrix W_U such that with probability $1 - o(1)$,

$$\|\hat{U} - UW_U\|_{2,\infty} \lesssim \frac{\log(n)}{n\rho_n}.$$

This theorem is particularly nice in that it allows for correlated graphs. I think there should be a factor of $\sqrt{1-\nu}$ on the right hand side if careful accounting of correlation is done, but that wasn't really the point of the result. The result also emphasizes how more refined analysis than the one leading to the proof of Theorem 4.2 can yield possibly better results. To be clear how this result is an improvement of Theorem 4.2 for this setting, in the example above, we have

$$\|U\|_{2,\infty} \lesssim \frac{1}{\sqrt{n}}$$
$$\lambda_r = \lambda_2 = \frac{n\rho_n(a-b)}{2} \gtrsim n\rho_n$$

so that the result reads

$$\|\hat{U} - UW_U\|_{2,\infty} \lesssim \frac{\|E\|_\infty}{n^{3/2}\rho_n},$$

which will not be very tight. Actually, using the fact that $\|\cdot\|_\infty \leq \sqrt{n}\|\cdot\|$, one can bound this by n, yielding

$$\|\hat{U} - UW_U\|_{2,\infty} \lesssim \frac{1}{\sqrt{n}\rho_n},$$

which is effectively useless if $n\rho_n \asymp \log^k(n)$ for any $k > 4$.

Finally, I think it would be remiss of me to completely ignore the high-dimensional covariance results since studying eigenvectors of symmetric matrices naturally leads to covariance estimation. The "spiked model" as in Theorem 1 is readily studyable using these techniques since it's got spiked eigenvalues and a notion of eigengap.

3 Discussion Points

- The proofs all require that $\lambda_r \geq 2\|E\|$, and though they allow for nonzero $r+1$-st eigenvalues, they are most readily available to be applied with zero $r+1$-st eigenvalues. How would these results change without
  - a spectral gap?
  - zero eigenvalues?
- How much does the grand decomposition affect the results?
- Classical multidimensional scaling (CMDS) is defined by taking the leading eigenvectors of $JXX^\top J$, where $J = I - \frac{1}{n}11^\top$ is the double-centering matrix. How might these results be applied to study CMDS? (Little et al. (2019) does this)
- How do these results illustrate the duality between high-dimensional covariance subspace estimation and classical multidimensional scaling?
- The proof of Theorem 4.7 uses Hoeffding's inequality to bound the term $EU\Lambda^{-1}$. How might other tools and techniques be leveraged to refine these results further?
- How could one use these decomposition techniques to calculate $\mathbb{E}(\hat{U})$? Is it possible?
- How would these results translate to studying scaled eigenvectors?
- Other entrywise norms?
- terms like $x^\top(\hat{U}\hat{U}^\top - UU^\top)y$? (hint: recall the "add and subtract of W_* to see that:

$$\begin{aligned}\|\hat{U} - UU^\top\hat{U}\|_{2,\infty} &= \|\hat{U}\hat{U}^\top - UU^\top\hat{U}\hat{U}^\top\|_{2,\infty} \\ &\leq \|\hat{U}\hat{U}^\top - UU^\top\|_{2,\infty} \\ &= e_i^\top(\hat{U}\hat{U}^\top - UU^\top)x\end{aligned}$$

  for some i and some x such that $\|x\| = 1$.

References

E. Abbe, J. Fan, K. Wang, and Y. Zhong. Entrywise Eigenvector Analysis of Random Matrices with Low Expected Rank. 2017. arXiv: 1709.09565.

T. T. Cai and A. Zhang. Rate-optimal perturbation bounds for singular subspaces with applications to high-dimensional statistics. *The Annals of Statistics*, 46(1):60–89, Feb. 2018. ISSN 0090-5364, 2168-8966. doi: 10.1214/17-AOS1541. URL https://projecteuclid.org/euclid.aos/1519268424.

J. Cape, M. Tang, and C. E. Priebe. Signal-plus-noise matrix models: eigenvector deviations and fluctuations. *Biometrika*, 106(1):243–250, Mar. 2019a. ISSN 0006-3444. doi: 10.1093/biomet/asy070. URL https://academic.oup.com/biomet/article/106/1/243/5280315.

J. Cape, M. Tang, and C. E. Priebe. The two-to-infinity norm and singular subspace geometry with applications to high-dimensional statistics. *The Annals of Statistics*, 47(5):2405–2439, Oct. 2019b. ISSN 0090-5364, 2168-8966. doi: 10.1214/18-AOS1752. URL https://projecteuclid.org/euclid.aos/1564797852.

L. Lei. Unified $\ell_{2\rightarrow\infty}$ Eigenspace Perturbation Theory for Symmetric Random Matrices. *arXiv:1909.04798 [math, stat]*, Sept. 2019. URL http://arxiv.org/abs/1909.04798. arXiv: 1909.04798.

A. Little, Y. Xie, and Q. Sun. An Analysis of Classical Multidimensional Scaling. *arXiv:1812.11954 [math, stat]*, Jan. 2019. URL http://arxiv.org/abs/1812.11954. arXiv: 1812.11954.

Z. Zhu, T. Wang, and R. J. Samworth. High-dimensional principal component analysis with heterogeneous missingness. June 2019. URL https://arxiv.org/abs/1906.12125v1.